DOCUMENT RESUME

ED 420 665 TM 028 339

AUTHOR	Stevens, David; McGuire, Michael
TITLE	Unit Record Versus Aggregate Data: Perspectives on Postsecondary Education, Data Collection, Retention, and Release. Final Report of the Working Group.
INSTITUTION	National Postsecondary Education Cooperative.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-98-280
PUB DATE	1998-04-00
NOTE	44p.; Federal
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Records; *Accountability; Data Analysis; *Data Collection; *Higher Education; *Information Dissemination; Institutional Characteristics; *Student Records
IDENTIFIERS	*Aggregation (Data)

ABSTRACT

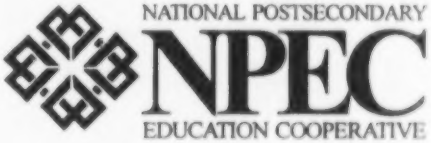
The Steering Committee of the National Postsecondary Education Cooperative established a Unit Record versus Aggregate Data Working Group in 1996 to identify and assess data sharing methods that are needed to advance the understanding of how postsecondary students are served, and with what outcomes. Most mandated record sharing has been through aggregate data methodologies, but some members of the Working Group advocate more widespread development of unit record methodologies. This paper adopts a practical approach to the unit record versus aggregate data topic. It does not argue the universal superiority of unit records over aggregated data, or vice versa. Rather, it examines and evaluates both approaches and the issues that will need to be addressed. After an overview of aggregate data, which refers to data collected at the institutional rather than the unit (student) record level, chapter 2 identifies three stages of data handling that usually involve different responsible parties, legal stipulations, and funding streams: collection, sharing, and release to users. Chapter 3 discusses new approaches, and chapter 4 considers the reporting and sharing process. Chapter 5 presents some recommendations for continued work by the Working Group, with a focus on potential linkages of unit record data to assist postsecondary entities in the development of statistics that will meet current demands for accountability. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 420 665

Unit Record Versus Aggregate Data: Perspectives on Postsecondary Education Data Collection, Retention, and Release

Final Report of the National Postsecondary Education Cooperative Working Group on Unit Record Data Versus Aggregate Data



TM028339

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Unit Record Versus Aggregate Data: Perspectives on Postsecondary Education Data Collection, Retention, and Release

Final Report of the National Postsecondary Education Cooperative Working Group on Unit Record Data Versus Aggregate Data

Prepared for the Council of the National Postsecondary Education Cooperative (NPEC) and its Working Group on Unit Record Data Versus Aggregate Data by David Stevens, Executive Director, The Jacob France Center, Merrick School of Business, University of Baltimore, Baltimore, MD, and Michael McGuire, Director of Research, Georgetown University, Washington, DC, under the sponsorship of the National Center for Education Statistics (NCES), U.S. Department of Education.

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Ricky T. Takai
Acting Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr.
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue NW
Washington, DC 20208–5574

April 1998

The NCES World Wide Web Home Page is
http://nces.ed.gov

Suggested Citation
U.S. Department of Education. National Center for Education Statistics. *Unit Record Versus Aggregate Data: Perspectives on Postsecondary Education Data Collection, Retention, and Release*, NCES 98–280, prepared by David Stevens and Michael McGuire for the Council of the National Postsecondary Education Cooperative Working Group on Unit Record Data Versus Aggregate Data. Washington, DC: 1998.

Contact:
Nancy Schantz
(202) 219–1590

MEMBERS OF THE UNIT RECORD DATA VERSUS AGGREGATE DATA WORKING GROUP

Jay Pfeiffer, Director, Office of Workforce, Education, and Outcome Information Services, Florida Department of Education (Chairperson)
Gary Barnes, Vice President for Program Assessment and Public Service, University of North Carolina General Administration
Mark Chisholm, Director, Office of Institutional Research, University of New Mexico
Thomas McCord, Director, Office of Research and Information Systems, New York State Education Department
Michael McGuire, Director of Research, Georgetown University
Marian Pagano, Associate Provost, Columbia University
Richard Voorhees, Associate Vice President for Educational Support Services, Colorado Community College and Occupational System
Jeff Weber, Manager, Information and Research, Indiana Commission for Higher Education

Consultants to the Working Group

David Stevens, Executive Director, The Jacob France Center, Merrick School of Business, University of Baltimore
Michael McGuire, Director of Research, Georgetown University

NCES Staff

Brenda Albright, Consultant to NPEC
Roslyn Korb, Program Director, Postsecondary and Library Cooperative Programs
Gerald Malitz, NCES Liaison
Joseph Rose, NCES Liaison
Nancy Schantz, NPEC Project Director
Robert Wallhaus, Consultant to NPEC

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION 1
Purpose 1
Scope and Direction 1

CHAPTER II

ISSUES FOR AGGREGATE AND UNIT RECORD APPROACHES 3
Aggregate Data Systems: An Overview 3
Collection 4
Sharing of Information 6
Confidentiality 7
Privacy Versus Confidentiality 7
Waiver of Right to Privacy 7
Informed Consent 8
Burdens and Benefits 9
Release Versus Disclosure 9
Action Needed 9
Flexibility 10
Definitional Clarity 10
Data Quality 11
Data Retention 12
Access to Retained Data 12
Accuracy of Data 13
Sample Versus Universe 14
Release to Users 16

CHAPTER III

NEW APPROACHES 19
Technology Factors 19
Program Delivery Factors 22

CHAPTER IV

REPORTING AND SHARING PROCESSES 25
Perspectives 25
Vertical and Horizontal Features and FETPIP 26
FETPIP's Postsecondary Features 27
Challenges Faced in the FETPIP Program 28
A National, but Voluntary Approach 29
A Bottom-Up and Voluntary Approach 31
A Within-State, Self-Contained Approach 32
The University of North Carolina General Administration 32
University of Maryland System 33
Interstate Data Sharing 34
Multi-purpose Accountability Systems 34
The National Student Loan Clearinghouse 35

CHAPTER V

RECOMMENDATIONS 39

Endnotes 41

CHAPTER I
INTRODUCTION

Purpose

The Steering Committee of the National Postsecondary Education Cooperative (NPEC) established a Unit Record Versus Aggregate Data Working Group in 1996 to identify and assess data sharing methods that are needed to advance our understanding of when, where, and how postsecondary students are served, and with what consequences. Specifically, the Working Group's charge was to: (1) contrast and evaluate benefits and limitations of unit record level reporting versus aggregate reporting; (2) identify and analyze factors that make unit record level reporting an issue, including confidentiality, flexibility with changing requirements and definitions, costs, and burdens; (3) recommend new approaches for collecting, maintaining, and sharing data taking into account technology, program delivery, and other changes affecting postsecondary education; (4) document and evaluate prominent unit record level and aggregate reporting processes and practices; and (5) document and evaluate record sharing practices at the unit level as well as the aggregate level between institutions, governing boards, and state and Federal governments.

Scope and Direction

The Working Group's title -- Unit Record Versus Aggregate Data -- suggests that the subject matter is contested ground. Some members of the group started from a premise that the status quo for most mandated data sharing -- aggregate data methodologies -- is satisfactory until proven otherwise. Other Working Group members detected inadequacies in the status quo and advocated the more widespread development of unit record methodologies. These divergent views did not prevent the Working Group members from concurring that a balanced treatment of the "contrast and evaluate" charges from the Steering Committee was needed, an approach that gives what we will call "fair treatment" to both unit record and aggregate data.

This fair treatment directive from the Working Group members to the consultants could lead some readers to assume that the paper will begin or end with a summary table of benefits and limitations under unit record and aggregate data headings. No such table appears here. The Working Group members felt that a table of benefits and limitations would provide a solid basis for further discussion only if a particular use, or narrow class of uses, of the data had been isolated for exclusive attention. Constraining the scope of the paper

in this way at the outset would be counterproductive. Instead, a more inclusive approach has been adopted to stimulate further discussion among readers who have diverse values, different educational and work experiences, and varied affiliations and responsibilities. The reader will find that the potential benefits and limitations are discussed throughout the document.

This paper adopts a practical approach to the unit record versus aggregate data topic. It does not advocate the universal superiority of unit records over aggregated data, or vice versa. Rather, it examines and evaluates both approaches and the issues that will need to be addressed should one methodology or the other be selected for a given research or administrative (e.g., student tracking) application. While fair treatment exemplifies the philosophical approach of the paper, the bulk of the discussion will focus on unit records for two reasons: (1) it is an emerging methodology, relatively speaking, whose use is presently less widespread than aggregate approaches, and (2) its use entails a host of procedural, technological, and ethical considerations that are absent from or at least less pronounced in aggregate methodologies.

The paper continues in Chapter II, after a brief overview of aggregated data, by identifying three stages of data handling that usually involve different responsible parties, legal stipulations, and funding streams: (1) collection, (2) sharing, and (3) release to users. This separation into three components builds on a basic fact-of-life about today's postsecondary education data elements -- most are collected initially as a unit record. Aggregation typically occurs when these unit record data elements are shared with some other person or organization, internal or external. Re-aggregation may then occur when one or more parties who have received the shared data prepare it for release to end users.

CHAPTER II

ISSUES FOR AGGREGATE AND UNIT RECORD APPROACHES

For the reasons noted, the majority of this paper focuses on issues pertaining primarily to unit record methodologies. To establish a fuller context for that discussion and to accomplish the goal of balanced treatment, however, this section begins with a brief overview of the issues associated with aggregated data.

Aggregate Data Systems: An Overview

Aggregate data refers to data collected at the institutional rather than unit (student) record level, typically through surveys. Examples include the Integrated Postsecondary Education Data System (IPEDS) surveys, the faculty compensation survey conducted annually by the American Association of University Professors (AAUP), and many others. While technically each institution might be considered a unit in these studies, since that is the level of aggregation for data collection and at least one type of reporting, for the purposes of this paper the term "unit record" is restricted to instances in which the most basic level of aggregation (hence, the unit) is the student. This is an important distinction, because some of the issues that will be addressed later for unit records are applicable to aggregated data as well, but in different ways. For example, confidentiality is also an issue for release of aggregated data at the institutional level when institutions are identified by name, but this is different in many ways from the confidentiality issues involved in unit record approaches.

Aggregated data provide summary information on many topics at the institutional level or at higher levels of aggregation. They are driven by common data element definitions and instructions for providing the data. They provide some of the same information (at institutional or higher levels of aggregation), and suffer from some of the same limitations (e.g., vulnerability to changing data definitions), as unit record approaches. However, aggregate data methodologies carry the burden of two additional limitations that have led increasingly to the advocacy and use of unit record approaches: (1) data initially collected at the institutional level cannot be used for lower levels of aggregation, for example, the tracking of individual students over time and across institutions/activities, and (2) aggregate data systems lack the flexibility to examine relationships among variables and to re-aggregate data, should reporting needs change. For aggregate systems, design choices restrict later analytical options to a far greater extent than is the case for unit record systems.

Aggregated data continue to satisfy many needs in postsecondary education. Institutions presently devote considerable effort to completing the many surveys that collect aggregate data, and they reap significant benefit from the studies performed within an aggregate data framework. For those information needs that cannot be met by aggregate systems, unit record systems have been proposed and in some cases developed. Our focus now shifts to those issues that affect the consideration and implementation of unit record approaches.

Collection

A comprehensive catalog of postsecondary databases would include a mix of unit record and aggregate data. Data elements that are collected today include a mix of information that is sought by the collecting entity itself, typically a college or university, in support of its self-defined missions, and data that are collected only because of an external requirement to do so. Some of the data elements reflect the referent institution's own management decisions and culture, each of which may change over time. The mandatory data elements are traceable to such external parties as governing boards, state authorities, Federal government entities, financial institutions, and accreditation bodies.

This section takes as its point of departure the central feature of the Working Group's overall contrast and evaluate charge -- identifying and assessing data that are needed to advance our understanding of when, where, and how postsecondary students are served and with what consequences. No attention is given to transactions within and between postsecondary institutions that have no, or only a tangential, connection to student flows into, through, and out of postsecondary education. Neither is attention given to other types of data collected in unit record format, such as employee or accounting records, although many of the same principles would be applicable to those undertakings as well.

A student's application to enroll, annual financial aid request, term-by-term registration and progress, and certification are usually documented in a unit-record context. There are often multiple storage locations within an institution for these different components of a student's transactions. This scattering of unit record data elements within an institution pales next to the number of unit record nodes that would have to be tapped to describe the odyssey of many high school students through a lifetime of postsecondary education transactions.

Today's data collection and retention practices will either enable or constrain tomorrow's uses of the data and their value in these future applications. A large and growing share of unit record data are collected in an electronic medium. Caution should be exercised in allowing current use of electronic unit record data collection to become an implicit criterion for including a student or institution in a referent population. Practical considerations, such as time and budget constraints, may require a reluctant adoption of this approach. If so, the resulting loss of information should be revealed to interested parties, so they have a better understanding of the limitations of the data that are collected. Moreover, the capability to carry out electronic data collection may be an appropriate explicit criterion for an institution's participation in a uniform unit record data collection to encourage them to acquire the capability using their own resources or to justify a request for external funds.

The distinction between collected and selectively recorded recognizes the common practice of only recording electronically some of the original data elements that appear on a paper source document. The extra electronic data-entry step introduces three new issues: (1) recording errors that depend on the quality-control steps that are taken, which vary greatly across institutions; (2) opportunities for others to impose their own judgment about what data elements to enter; and (3) differences across institutions in paper or electronic (e.g., data entry screen) source documents that will influence comparability in later sharing and release.

Isolation of the collection stage highlights one important aspect of the contrast and evaluate responsibility assigned to the Working Group -- unit record collection enables more flexible subsequent sharing and release to users than aggregate data collection. As mentioned previously, information that is collected in aggregate form can never be disaggregated in the future. Data that are collected in unit record form can be aggregated later, and then re-aggregated again, assuming that the unit record data have been retained.

At least two other important considerations apply to unit record, and to a lesser degree aggregate, data collection:

(1) When the original data collection initiative comes from an outside source, the probability of gaining voluntary institutional participation will increase if sufficient value is seen in the potential use of the new information for the institution's own purposes.

(2) An institution's opportunity for voluntary participation in an external party's data collection initiative is often assessed relative to the perceived threat that is posed by placing new detailed information in the hands of individuals or organizations whose motives are viewed with skepticism or even fear.

This tension between perceived opportunity on the one hand and threat on the other highlights a critical contextual dynamic operating within postsecondary education and society in general: distrust of authority and reluctance to cooperate voluntarily in initiatives sponsored by that authority. This dynamic operates at the micro level -- for example, individual students' distrust of their college's administration -- and at the macro level -- institutions' distrust of the media and state and Federal agencies. Well-publicized if not widely practiced abuses of electronic information, with resulting harassment or exploitation of private citizens, provide an even broader backdrop against which this distrust has developed. Those who plan and implement unit record data systems should anticipate resistance to access to unit record information and should be prepared to demonstrate legitimate and compelling benefits -- and guarantees against abuses -- to gain the cooperation of individuals and institutions.

Different people, institutions, funding streams, and purposes are involved in the handling of postsecondary education data collection, sharing, and release. This diversity offers a plausible explanation for the observed continuum of opinion that extends from vigorous opposition to unit record data sharing to aggressive advocacy of this approach. Many of those who bear the costs and risks of data collection are far removed from the payoff on this investment. Bridging this gap and providing real value, or at least an explanation of the larger value, to those involved in data collection is an important ingredient of successful and responsible project management.

Sharing of Information

The second charge to the Working Group was to "identify and analyze factors that make unit record reporting an issue, including confidentiality, flexibility with changing requirements and definitions, costs, and burdens." Selected factors that make unit record reporting an issue are covered next.

Confidentiality

The topics covered here are: (1) privacy versus confidentiality, (2) waiver of right to privacy, (3) informed consent, (4) burdens and benefits, (5) release versus disclosure, and (6) action needed.

Privacy Versus Confidentiality

Privacy and confidentiality are not identical. Privacy refers to an individual's right to withhold information, that is, not to divulge information to anyone else. Confidentiality refers to the handling of information that has been obtained by a second party. The unit record information that exists in the electronic and paper files of postsecondary education institutions is there because students or others implicitly or explicitly waived their right to privacy. One of the common explanations for missing data elements in unit record databases, such as a student's social security number or ethnicity, is that the right to privacy was selectively invoked by the individual student to justify withholding that information.

Waiver of Right to Privacy

A person's willingness to waive the right to privacy is often driven by the unknown or feared consequences of that action. Those who observe a growing distrust between students and postsecondary institutions around the issue of confidentiality express concern about how this attitude affects reporting accuracy. This in turn influences how the potential benefits to be derived from unit record, or any other, data sharing initiatives are evaluated.

Typically, anyone affiliated with a postsecondary institution is informed that certain individually descriptive data elements will be defined as directory information, which means that these data can be released to the public. Other information is described as confidential, which means that it will not be released in a way that reveals the identity of the person who provided it, or to whom it pertains in the case of transcript and some financial data. Different Federal, state, and perhaps even institutional rules cover particular classes of affiliation, such as student or employee.

Informed Consent

A basic confidentiality issue is whether a person has an absolute right to be given an opportunity to provide or withhold informed consent (i.e., to waive or retain voluntarily the right to privacy) for each and every proposed use of unit record information. If so, a practical question remains: How should this be carried out? There is an extensive conceptual literature, documentation of actual practices, and judicial record on this issue. A fundamental theme is whether, and under what circumstances, implied consent can be assumed. Another relevant topic is when and where an official record of a person's consent should be maintained and for how long.

This was one of many issues before the Working Group. Legal and practical differences exist between informed consent stipulations that require an individual's consent and those that allow someone else to act on the individual's behalf. For example, the unemployment insurance units in state employment security agencies collect unit record information about employee earnings from employers who are required to provide this information by state unemployment compensation laws. Individual employees do not have an opportunity to ask their employer to withhold this information. Their consent is implied by the acceptance of, and persistence in, the job. Anyone who inquires is usually told that the information provided may be used for statistical purposes other than the immediate program application, but that the information will remain confidential and will not be released to the public in a way that directly or indirectly reveals a the employee's identity. The Bureau of Labor Statistics provides the same assurance of anonymity to reporting businesses. Critics of the statistical purposes warning include those who allege that informed consent is meaningless in such a vague context of possible future uses of the information.

A different approach to getting informed consent has been proposed by one of our Working Group colleagues. Each student has a selfish interest in the success of the postsecondary institution attended. Given that an efficient and effective institution is more likely to prosper than a poorly managed one, a student can expect to share in the benefits that flow from an institutional performance advantage. This benefit might accrue in the form of a smaller tuition increase or as a career advancement bonus associated with the rising stature of one's alma mater. This self-interest motive is a way to encourage students to grant voluntary informed consent for the use of unit record data. It remains to be seen whether perceived self-interest will be strong enough to override distrust of the institution or others in many cases.

Recent anecdotal evidence of mishandling of confidential records in noneducation settings has increased public concern about the willingness and ability of those who promise confidentiality to honor that pledge, adversely affecting the public's willingness to grant consent. Furthermore, reluctance to offer informed consent often has cultural, religious, or other value-based origins which are not evenly distributed across the postsecondary education community, so important limitations in the interpretation of unit record databases are likely to arise.

Burdens and Benefits

Those who bear the burden and cost of collecting data are usually not the same people or institutions who derive the value from this investment when information is released to users. A similar difference of interests and motives is found when the confidentiality issue is investigated. Those who are charged with convincing people to consent to permitting use of unit record information are usually far removed from those who will derive the benefit from their success.

Release Versus Disclosure

It is important to distinguish between the release of information to an external party and disclosure of an individual or institution's identity. The latter is often prohibited, but release of unit record information is rarely barred. Today's technologies offer an assortment of ways to accommodate release of unit record data without jeopardizing restrictions on identity disclosure. The section on technology factors in Chapter III covers some aspects of this issue. Legal opinions that distinguish between release and disclosure are now available for use by those who seek access to confidential unit records but have no intention of disclosing the identities of the people and businesses or institutions that are represented in the database.

Action Needed

Some observers advocate legislative action to amend Federal confidentiality laws to facilitate the sharing of unit record data. Other observers consider such action to be a last-resort step which may not be needed to accomplish high-priority data sharing initiatives. There are a growing number of successful data sharing activities that offer guidance about how to succeed within today's legal and regulatory environment.

Flexibility

The rationale for unit record data sharing that one hears most often is flexibility. Even a cursory reading of postsecondary education history makes it abundantly clear that Federal and state laws, governance structures, and management fads are subject to repeated, and not necessarily predictable, changes. Yesterday's data element definitions and aggregation rules may not satisfy today's needs, and refinements to meet today's requirements are unlikely to suffice in tomorrow's environment.

Four issues are of particular importance when flexibility is valued: (1) definitional clarity, (2) data quality, (3) data retention, and (4) access. Each of these four attributes is necessary for the value of flexibility to be maximized. Definitional clarity, data quality, and access pertain to both current and future uses of data that have been collected. The data retention topic affects only future uses.

Definitional clarity and data quality influence flexibility because each use can be assigned a place on a continuum based on the combined importance of two factors: (1) definitional precision and (2) quality assurance that the intended standard of precision has been met. The higher the levels of definitional precision and quality assurance, the greater the flexibility in using the data for varied purposes. Some readers may think of definitional clarity and data quality as data collection rather than flexibility topics. Treatment of these issues was deferred until now to emphasize the point that value arises from the use of data, not from the collection of data.

Definitional Clarity

Changes in data element definitions, such as the evolution of ethnicity designations in recent years, typically occur at the unit record data collection point. Some redesigns of data collection instruments make a conscious, but not always successful, effort to ensure a precise mapping between old and new definitions. Other redesigns pay no attention to this issue, which can limit an analyst's ability to interpret observed differences between base-period and end-period data points.

Much of the postsecondary education information that one might seek for a wide range of uses is now collected at the unit record level somewhere by someone. Those who participate in today's collection of information may not know what subsequent analysis, sharing, and release will occur. This knowledge gap would be expected to have data quality implications, which are described next.

Data Quality

Lack of communication between data collectors and users can result in a loss of unit record accuracy. The use of a student's social security number as a unit record identifier illustrates this point. A widespread practice in the past has been to substitute a pseudo nine-digit number series for a missing social security number on a student's record when this field is used as the official student identifier. For many internal purposes this substitution practice does not matter. However, accuracy may be critical when this record is placed in a data sharing context that involves merging and tracking, especially when the value of the identifier itself changes over time. In general, the importance of accuracy should be conveyed back to the collection point and steps should be taken to ensure that changing values are updated in a consistent and timely manner.

A second source of diminished accuracy arises from the actions of those who know what might be done with the records but then cannot or do not take the necessary precautions to ensure that the desired level of data quality is reached. Members of the institutional research community are routinely faced with vexing challenges to their own professional standards when unit records are shared, aggregated, and released in ways that do not always reflect known deficiencies in the underlying data. This is one reason to endorse aggregation in some applications when more detailed unit records cannot be collected at an acceptable level of accuracy. It is doubtful whether aggregation solves the accuracy problem; more often than not, it merely sweeps measurement error under the rug. This set of circumstances also raises a secondary series of questions: Who determines the acceptable level of accuracy? Is this threshold enforced; and, if so how and by whom? What sanction is imposed for not ensuring that the threshold quality standard is met or exceeded?

The authors' stated reluctance to offer a generic summary table of the pros and cons of unit record versus aggregate data is grounded in part in the points made in the previous two paragraphs. Those who are responsible for the collection of unit record information may not know or care about the use-specific quality standards that apply in particular data sharing and release situations. Stated another way, the issue is rarely whether available data are perfect. The pertinent question is usually: "Are the data good enough for the intended purpose?" The three-stage sequence of collection, sharing, and release helps us to see why this is a difficult question to answer. If those who are responsible for quality control at the collection point do not know the full range of intended uses, then they cannot offer expert counsel on the quality issue. Similarly, if those who bundled unit records in shared-use settings are isolated from those who collected the original data elements, they too may be unable to address the question of adequate quality.

Data Retention

This is a complex but essential issue to consider under the umbrella of flexibility. Flexibility is enhanced by definitional clarity and enforcement of a uniform threshold of data quality, but these conditions only affect today's sharing and release if data are not retained for future use. Two pertinent questions are: Who has an interest in data retention? and How will retention be accomplished? One answer to the interest question is: usually someone, or some organization, other than the original data collector. A reasonable response to the how question is: relatively easily, given today's technologies, but not without important political and administrative hurdles.

Data retention is more or less neutral with respect to the unit record versus aggregate data topic, but the ability to perform customized bundling of unit record data is a dependable way to increase the total payoff on any investment in data retention. It is important to acknowledge the approximate cost-neutrality of data storage and retrieval in an electronic medium and to assume the willingness and ability of the data retention intermediary to comply with the letter and spirit of confidentiality stipulations. If either or both of these assumptions are dismissed, an aggregate approach might be deemed superior, that is, safer and/or more cost effective.

Access to Retained Data

Again, the third stage in the collect, share, and release sequence is crucial because the value of data sharing only emerges from use. The intermediate step of data retention is a necessary but not sufficient condition for realizing a future payoff on today's investment in collection.

Answers to the following questions will provide a starting point for deciding how much flexibility is enough, which is one basic criterion for approaching the unit record versus aggregate data issue in a use-specific context:

(1) What statutory, other mandated, and discretionary users of the data elements can be identified with relative certainty? The answer to this question identifies the population of users to whom data must be released, now or in the foreseeable future.

(2) What are the minimum and maximum levels of aggregation that each of the current users will tolerate? This places a floor under and a ceiling over how much aggregation can occur.

(3) Would accommodating the needs of any one of the potential future users change the aggregation range? If so, one may have to decide whether to make the necessary investment to lower the floor in the present to satisfy a future need.

Accuracy of Data

An overarching principle about accuracy in the context of unit records versus aggregate data is that the degree of accuracy that is required cannot be defined without reference to the intended use of the data. A high level of accuracy is costly to achieve. This may not be needed, or if needed, may not be possible to attain.

The more steps that occur in moving data from the point of collection to the eventual data user (typically, an analyst or tracker) and the fewer the quality-control mechanisms in place, the less likely that user is to understand the full range of measurement errors that might have emerged during this passage. This issue is particularly acute when the original data collector receives neither reward nor sanction for maintaining known quality standards. Each of us has heard the common assertion that "the data are just for a report that we have to submit," which is one confidence-draining way of responding to the question about quality.

Up to this point, a one-way flow of data from collection, through sharing, to release and use has been implied. This approach tells only part of the story. Equally, and some would say, more important are cases in which a data collector wants to merge internal data with other information that is controlled by one or more external parties. In such instances, the accuracy of the records that are already maintained internally may not be anyone else's concern. This would not be true, of course, if reciprocity is expected. In that case, the members of the data sharing group would normally be expected to agree upon common rules of data availability and accuracy.

For example, let us assume that a public university system wants to know whether and where a particular population of former students are working. Some, but not all, public university systems have an institutional research capability to prepare an electronic file of student social security numbers to be submitted through the Information Technology Support Center's distributed database to each participating state employment security agency. This assumes that a student's social security number is either used as the official student identifier, and is accurate, or that accurate social security number information is collected and maintained using some other rationale.

In this example, each cooperating state employment security agency would attempt to match the transmitted social security numbers with its administrative records of employment and earnings and then return agreed-upon data elements back through the distributed database system to the university. The university has not shared or released any of its own administrative data except the social security numbers. However, other parties may seek to use the distributed database capability to find out whether designated individuals have enrolled elsewhere in the university system and, if so, to receive agreed-upon data about those students through the distributed database system. The university system retains total control over and responsibility for accuracy regarding its own administrative records, but the system is expected to meet agreed-upon accuracy standards when its records are shared with other agencies.

Sample Versus Universe

For the purposes of this discussion, universe means all members of a designated population and sample means any defined subset of a population. Some of the historical reasons for favoring the collection and sharing of sample data are no longer relevant to a comparison of the pros and cons of unit record versus aggregate data. When all steps are electronic, advances in the electronic collection, storage, transmission, and manipulation of data have effectively eliminated cost differences between sample and universe as a deciding factor. In fact, sampling could be more expensive than universe inclusion, depending upon the techniques and technology used to identify and draw the sample.

The continued presence of a mixture of electronic and other ways of collecting and storing information introduces an equity issue -- the nature and extent of costs will differ among those who use different collection, storage, retrieval, and analysis methods. The Working Group has made no attempt to estimate the range of such costs, so there has been no discussion of possible responses to an awareness of these cost differences. The sample versus universe debate would be advanced by having accurate estimates of these costs available to those who participate in such dialogues. Related cost information about the Department of Labor's proposed distributed database, for example, should be available soon.

Many who assert the sufficiency of the sample approach usually have a well-defined, singular purpose in mind, such as a survey of the postsecondary education plans of the current year's high school seniors. Others who advocate the retention of universe data are often looking beyond a single immediate use of the information to realize additional value from the data in the future. Acknowledgment of these assumptions

helps to identify the differing motives of the respective parties and to devise a strategy to link the decision of sample or universe to the goals of the study. Some who advocate the collection of sample data do so as a conscious step toward limiting the future range of uses that can be made of the information. Similarly, those who seek the retention of universe data may not have invested the time and effort to define the practical limits on data value that would result if a sample was accepted. This describes the stance of many members of the postsecondary education research community who focus on data value for a particular use without giving much thought to the costs incurred in data collection, retention, and sharing.

It is also important to understand the causes and consequences of attrition over time from a beginning sample, and what happens when a future data need cannot be satisfied because the original sample was not designed with that need in mind. It is usually possible to speculate intelligently about expected attrition of a base-period population over time based on previous attrition histories in related studies. It is much more difficult to persuade skeptics that a particular future use of the data might arise, which requires modification of a sample design that does not take this use into account. An extreme modification would be to include an entire population. In fact, this modification may not be extreme in its impact on collection cost or burden, depending on the way in which the desired data elements are acquired.

While there are examples of recent national sample designs for the collection of postsecondary education data by survey methods, most of the sample versus universe issues arise in the sharing phase. Again, most data that are of interest for many purposes are already collected in unit record form somewhere. If quality differences exceed a user-defined tolerance level, then the sample versus universe issue is moot because whatever information is available is not usable.

One possible reason to favor sample data over the retention and sharing of universe data arises from widespread mistrust about the willingness and ability of those who manage databases to protect the confidentiality of the records that are maintained. There may be no logical distinction between the vulnerability of sample records and universe records, but many people appear to be comforted by the knowledge that only some members of a designated population might be at risk of disclosure of information about them. Another reason to favor sample over universe data retention is to preclude the possibility of allowing others to use the database for law or other enforcement purposes. The legal boundaries for permitting or coercing such extraneous but powerfully motivated uses of data are in a state of flux.

Release to Users

Stage three of the collect, share, and release sequence of postsecondary education data handling is where the payoff on the investment in data collection and sharing appears. Higher payoff is often achieved with a broader client base (i.e., more end users of the data). Customized bundling of data elements to satisfy different user needs is one way to maximize value in a data sharing environment. The opportunity to bundle is limited by aggregation.

The interest and ability of institutional research professionals to conduct sophisticated multi-variate analyses using postsecondary education data elements is another reason to consider the collection and sharing of unit record information. Aggregate data are typically summarized in a univariate format; while possible to produce, multi-variate tabulations of aggregate data are cumbersome to interpret and to submit to statistical tests that are most appropriate at the unit record level.

A basic theme of the paper up to this point has been that most data elements are collected as unit records, which can be stored, shared, and released in multiple formats at a decreased cost, to produce higher potential benefits for end users. This functional approach ignores a basic and powerful principle of organizational behavior: Do not knowingly provide adversaries with weapons to attack you. Postsecondary education is not alone in trying to control the story that is told of its academic, social, and economic role. Control of the "spin" may well be lost once data are shared with others -- particularly the media and certain governmental entities -- whose credibility and motives are unknown, or known and feared. The culture of distrust referenced earlier should not be casually brushed aside.

On the other hand, the access and accountability questions that the public is posing to the postsecondary education community are legitimate and deserve a responsible, informed reply. The justification of both public and consumer spending on higher education needs to be addressed. Rational and effective public policy needs to be developed and, even though occurring within a political context, informed by accurate and timely data. Fear of negative spin should not be used as a shield to ward off public inquiry and, in any event, will be ineffective if so used.

One of the most important emerging factors in the accountability of postsecondary education is a growing ability to acquire and retain unit record data that will provide unprecedented insights into the relative importance of particular institutions, activities, and characteristics in a student's success, however defined.

Any student's achievements are a joint outcome of many forces. Awareness of this complexity and interdependence explains an expression of concern by many NPEC Council members -- that uncontrolled release of available, but oversimplified or otherwise lacking, outcomes information may distort rather than enhance the public's understanding of postsecondary education and any institution's accomplishments within this diverse community. In other words, the release of information can sometimes be seen as a cost instead of a benefit.

This point of view was discovered by the field researchers who prepared the NPEC briefing paper on *Student Outcomes From a Data Perspective*. That briefing paper states that "at one extreme, several respondents . . . believed that the state should stop collecting and disseminating occupational data since such data do little (in their opinion) to inform questions of institutional performance or effectiveness." In such cases, non-release of available information may be preferable to release of data that are too limited to ensure proper use or when the user is not deemed competent or trustworthy.

An ongoing challenge for postsecondary education leaders is to find a satisfactory approach that weighs the postsecondary education community's own need for better data against the external opportunities and threats that arise from the enhanced information sharing and release that unit records afford. This may be one of the common threads that weaves together the separate contributions of the 1996 NPEC working groups.

Chapter III
NEW APPROACHES

The third charge to the Working Group was to "recommend new approaches for collecting, maintaining, and exchanging data taking into account information technology, program delivery, and other changes affecting postsecondary education." This chapter combines additional treatment of some topics that were introduced earlier with new issues. The Technology Factors section identifies electronic communication opportunities that have redefined data collection practices and the links among collection, retention, and release. The *Program Delivery Factors* section begins from a premise that distance education and recurring adult participation in postsecondary education will continue to expand.

Technology Factors

Today's evolving postsecondary education environment in the United States is an ideal laboratory for examining the technological aspects of unit record versus aggregate data. The sensitivity of student enrollment decisions to pricing policies, the responsiveness of prospective students to various marketing strategies, patterns of student movement within and among institutions, re-enrollment phenomena, and the complex causal forces that underlie each of these issues are all ripe for study. Progress in gaining valuable insights about these and other topics is now more likely because of advances that continue to appear in the technology sphere. Software is readily available to transform prohibitive manual inspection tasks into routine electronic analysis of transactions data.

Transaction analysis can begin when a telephone, mail, fax, or electronic inquiry about an institution's offerings arrives from a given person and then continue through every subsequent interaction between the person and institution. Aggregation is necessary if an accurate personal identifier is not assigned and used consistently for tracking purposes. The proper use of personal identifiers has had a mixed track record to date, for a variety of reasons. This limitation may be a temporary circumstance that will fade over time as a constraint on creating and maintaining longitudinal unit records.

The goal of expert system software is to automate those aspects of human interaction that can be routinized in a dependable manner. This does not mean that mistakes are eliminated; it simply means that a

conscious effort is made to anticipate the nature of the errors and minimize them. The sophistication of software design continues to grow, which means that the power and user-friendliness of the resulting capability is improving.

Vendors now offer generic system designs that can be refined to satisfy a client's needs. For example, a postsecondary institution can install a multi-purpose communication system that automates such activities as responding to inquiries about particular course and program offerings, through registration and billing, to transmittal of grade information and full transcripts. When a first-time request for information is received from a touch-tone telephone or computer, the caller can be instructed to enter a Personal Identification Number (PIN) of a specified type, so any follow-up contact can be linked to this initial event. Then, an electronic record can be created and added to a database that keeps desired information about the inquiry and any response. The basic design of the system can be modified after transaction information becomes available, as glitches in the original decision tree are revealed and corrected.

The growing number of vendors who offer this type of service, and the expanding population of clients, does not reflect any national initiative or incentive. This means that cross-institutional compatibility of information collected has not been considered by the adopting institutions. The vendors have an incentive to encourage the sale of off-the-shelf or common designs to multiple customers, but the motive behind this is profit-seeking, not uniformity of the resulting transactions information.

There are many who would benefit from a better understanding of how people decide whether and where to seek information about postsecondary education opportunities, and what transpires after initial inquiries are made, because the social benefits that accompany informed choices complement the private benefits that accrue to those who make the decisions. Consistent use of a single identifier by each applicant across time and institutions would permit valuable transaction analysis to be carried out. For reasons described earlier in the paper, including the isolation of those who bear the data collection burden from much of the value that is extracted by others from that information, it is unlikely that widespread voluntary use of such an identifier will happen soon.

A practical way to move toward broader adoption of transaction analysis capabilities is to begin with clusters of institutions that have a common interest in voluntary cooperative data definitions and information collection, retention, and sharing practices. Obvious candidates include members of an association or consortium, institutions within a public postsecondary education system, and schools that have traditions of

competition for enrollees. Even then, use of a common identifier may need to be mandatory before the payoff to institutional participation can be demonstrated with actual data received from cooperating institutions.

Progress in electronic data sharing capabilities and practices is becoming more widely known as start-up efforts reach the third stage of actual data release. Florida's Education and Training Placement Information Program (FETPIP), described in Chapter IV of this paper, is a pioneering example. At least three other states (North Carolina, North Dakota, and Texas) have adopted major features of this comprehensive approach to routinizing the availability of employment and earnings information on behalf of participating postsecondary institutions. After years of investment in system design and development, a shared information system created through the leadership of Oregon's Employment Department has just reached the stage three "release of data" point. Each of these is a state-initiated effort. There are many other similar examples. This growth of state initiatives is consistent with a point made in the previous paragraph -- the presence of a common interest paired with a pervasive lowering of cost and trust hurdles is a promising signal of success.

Technology has made unit record data retention and sharing more feasible, and potentially safer from a confidentiality protection standpoint, than in the days of paper records stored in file cabinets. Reliable encryption software is now in routine use. Other recent innovations include fingerprint technologies that require a person to log into a restricted database by placing a finger on the computer screen, which is then scanned for confirmation and authorized access using a previously stored database of approved fingerprints. This means that every transaction using the confidential records is recorded, and the responsible person can be identified and dealt with as needed if unacceptable behavior ensues. Progress on these fronts translates into an ability to satisfy more information users. A larger population of beneficiaries creates an incentive for further technological breakthroughs motivated by expected profits.

The speed and extent of advancing data storage and processing capabilities, and falling costs incurred in these activities, have affected the balance of opportunities and threats associated with data sharing and use. On the other hand, new opportunities for collecting, retaining, sharing, and using unit record information also create an unprecedented possibility of system failure. There is a compelling need to accompany progress in taking advantage of what technology allows with a parallel, and hopefully forward-looking, dialogue about what untoward events might be unleashed by too hasty action. A delicate balancing act is required here: Those who tarry too long in demonstrating the value of their innovation may be trampled by others who are not as cautious.

Program Delivery Factors

Three features of postsecondary education in the United States today are changing the nation's estimates of the net benefits that flow from unit record versus aggregate data collection and sharing:

(1) The rapid growth of distance education;

(2) Higher, and allegedly increasing, transfer rates among traditional students and periodic re-enrollment of adults in both credit courses and other types of postsecondary offerings; and

(3) The tendency for many students to pursue a post-baccalaureate degree -- increasingly, in many fields, an advanced degree is either required or desired for career growth.

Aggregate units of analysis that sufficed yesterday, such as institution and program, remain useful for some applications today but are inadequate for others. Designation of an outcome as attributable to only the most recent educational experience has always been questionable, but the weight of evidence indicating that this should not be done has become increasingly compelling as more students enroll at multiple institutions. The concept of joint outcome that was introduced and defined earlier is relevant here.

Any outcome measure, whether it be of cognitive competence, interpersonal skill and adaptability, employment and earnings, or anything else, is a measure of cumulative achievements and shortcomings, unless an accurate recording of before-and-after levels has occurred at the beginning and end of a particular period of exposure to postsecondary education. This does not mean that all previous forces are of equal influence. It does mean that a conscious effort should be made to account for those factors that most observers agree can and should be accounted for.

The growth of distance education and recurring enrollment patterns has affected the "should" part of this deliberation. The technology and cost of data storage and processing has affected the "can" issue. Participation in distance education opportunities and the push-pull of incentives to enroll throughout one's adult years occur over time. This translates into a growing interest in maintaining longitudinal data sets. It is not sufficient to simply maintain data over a long period of time. The databases have to be consciously designed to interlock. Many of today's data collectors have made no apparent effort to facilitate someone else's ability to link data with one or more other databases. This is understandable for reasons that have been described earlier.

A new compact between data collectors and data users would enable evaluators and policy analysts to study postsecondary education as a continuing, longitudinal phenomenon. The delivery of postsecondary education in the United States is changing rapidly. Many of the transactions that make up this structural change are needlessly lost to analysis; we say 'needlessly,' because this unit record information is often collected in some electronic medium and then consciously discarded because it is not believed by its custodian to be of use anymore. Strategic decisions are compromised by this gap in understanding.

Advocates of unit record data argue that sensible and effective policies in many parts of the postsecondary education enterprise today require timely and accurate information that can only be assembled through the collection and sharing of unit record information with appropriate respect for institutional and personal confidentiality standards and the diversity of missions across postsecondary education. They do not argue that all unit records should be retained forever, nor that unit record information should always be shared for the greater good. However, new opportunities for realizing the value from retained unit record information have been demonstrated (see Chapter IV). Perceived threats can and should be acknowledged and answered in a deliberate manner. Incentives for voluntary participation in unit record projects, rather than externally mandated occasions, should be created.

CHAPTER IV
REPORTING AND SHARING PROCESSES

Perspectives

Postsecondary education data are shared daily. One basic issue before the Working Group was how to realize higher value from shared data. There are many ways to increase value from sharing, some examples of which are provided below. These should be treated as illustrative of the range of such activities that are already under way. Other similar data sharing activities could have been chosen for coverage here.

A construction metaphor captures the theme of this section. Architects understand the difference between load-bearing and non-load-bearing features. Non-load-bearing components can be removed without affecting the structural integrity of the overall design. Like architects, the designers of postsecondary information systems should clearly identify the essential components that cannot be sacrificed without jeopardizing the integrity of the systems themselves. Justification of discretionary features can then proceed based on other criteria.

Postsecondary information systems appear, evolve, and sometimes fade from view in quite different ways over varied lengths of time. These information systems often have both vertical and horizontal features. Systems that include dissimilar types of institutions have vertical features, and those that cover similar institutions have horizontal attributes. The information system maintained by Florida's Office of Workforce Education and Outcome Information Services serves below as an example of both vertical and horizontal features.

The process of designing and then building a postsecondary information system can be described in top-down or bottom-up terms. Information systems of the top-down type usually include comprehensive design features at the outset. The distributed Wage Record Information System (WRIS) being developed by the Unemployment Insurance Information Technology Support Center offers an example of a top-down information system. Bottom-up systems often reflect the fact that they were assembled using available data components that do not necessarily result in a complete and unduplicated system of records. The postsecondary information system being assembled by The Jacob France Center at the University of Baltimore illustrates this bottom-up approach.

Each combination of the vertical/horizontal and top-down/bottom-up design features can be segmented further into single-state/multiple-state/national coverage. The information system managed by the University of North Carolina General Administration, and another developed by the University of Maryland System, are single-state examples. Washington State's Board for Community and Technical Colleges has been a pioneer in the design and practice of interstate data sharing. The National Student Loan Clearinghouse is a recent entrant into the national coverage market.

Census Bureau data, National Center for Education Statistics databases, and the Luxembourg Income Study serve as models of top-down national information systems that contain unit record data elements. In each case, user access is restricted, and release of identifiable information is prohibited. This statement applies to each of the postsecondary information systems previously identified in this section -- none permits release to the public of information that identifies a current or former student.

Vertical and Horizontal Features and FETPIP

The comprehensive information system now managed by Florida's Office of Workforce Education and Outcome Information Services -- the Florida Education and Training Placement Information Program (FETPIP) -- provides a valuable case study of most of the issues raised in the previous subsection[1]:

- Today's comprehensive information system has evolved over more than a decade (more than two decades if pilot phases of component testing are included) of almost continuous refinement.
- The genesis of FETPIP can be traced to the same issues that have been debated by Working Group members: (1) public and legislative interest in improved accountability for investments in public education, (2) concern about the quality of information then available for performance measurement purposes, and (3) anticipation of level or declining public investment in data collection and information system maintenance.
- FETPIP's management team has maneuvered through potential minefields by mixing: (1) successful legislative initiatives, (2) a persistent broadening of the system's customer base and constituent services, and (3) unwavering attention to data confidentiality matters as waves of Federal and state law changes and orchestrated assaults on public opinion have affected the program's opportunities and practices.
- The database architecture, processing protocols, and management rules exemplify most of the feasible combinations of unit record collection, longitudinal maintenance, controlled access, and multi-faceted release with different aggregation features.

- The FETPIP offers comfort to those who loathe one-size-fits-all initiatives that pay little heed to the interests of the data providers. The system now offers well over one hundred customized reporting applications for a growing number of interested parties.

- The system's reputation has grown through careful assembly and distribution of reliable evidence that the public's investment in FETPIP has produced an attractive return-on-investment shared by diverse constituencies.

A basic lesson to be drawn from the FETPIP's first decade and current status is that a state can carry out almost any conceivable range of performance measurement provisions that might emerge from pending reauthorization of the Higher Education Act, the Adult Education Act, the Carl D. Perkins Vocational and Applied Technology Act, and possible related amendments to the Indian Education Act, the Job Training Partnership Act, and other complementary Federal laws.

FETPIP's Postsecondary Features

The database maintained by Florida's Office of Workforce Education and Outcome Information Services contains unit record data elements for all community college associate degree and vocational students; all postsecondary vocational students in district-managed schools; all state university system graduates; adult education students; selected private vocational schools, colleges, and universities; all Job Training Partnership Act programs and Project Independence (welfare jobs program) participants; and other smaller programs that sometimes involve postsecondary institutions. The total number of participant records, including but not limited to postsecondary students, that are included in a data processing cycle now exceeds 2 million records.

Records are linked electronically with such outcome coverage sources as the State Department of Education (for both public and private college and university enrollments in Florida), the State Department of Management Services (for Florida career service employment), the Florida Department of Labor and Employment Security (for Florida employment covered by the state's unemployment compensation law), the Department of Children and Families (for public assistance participation), the Department of Corrections (for incarceration/probation information), the U.S. Postal Service (for postal career service employment located anywhere), the U.S. Department of Defense (for military personnel stationed anywhere worldwide), and the U.S. Office of Personnel Management (for Federal civilian employees wherever they may be assigned).

Based on the records obtained from the Florida Department of Labor and Employment Security, approximately 25,000 Florida employers of former high school and postsecondary students and other trainees are surveyed to collect occupational assignment and county worksite of each member of the reference population. This is complemented by employer opinion surveys that are designed in cooperation with such groups as the Division of Community Colleges, the Board of Regents, the Postsecondary Education Planning Commission, and employer advisory bodies.

The FETPIP participates in a Vocational Education Performance-Based Incentive Funding initiative, which includes involvement in an Occupational Forecasting Conference and the identification of "fundable" placements. The Occupational Forecasting Conference is a multi-agency activity that results in an agreed-upon list of occupations that offer the most promise for successful career entry and sustainability for Florida's students. The fundable placements concept refers to a cooperative effort to identify those students who successfully enter occupations on the Conference list, so their institutions can be rewarded for acting on the Conference information. By July 1998, the Performance-Based Incentive Funding program is to be expanded to cover all adult general education and postsecondary vocational and associate of science programs in Florida as a result of 1997 legislative action.

Challenges Faced in the FETPIP Program

The FETPIP management team encounters and attempts to respond to challenges that are voiced by those who use the program's information on a voluntary or mandated basis.

- The data provided by Florida's Department of Labor and Employment Security and by Florida's Department of Management Services cover only Florida employment. Only Federal civilian employment, military personnel, and U.S. Postal Service employment beyond Florida's borders is included. The Wage Record Interchange System now being designed by the Unemployment Insurance Information Technology Support Center, which is covered under *A National, But Voluntary Approach*, responds to this limitation.

- The earnings figures provided by Florida employers to the Department of Labor and Employment Security are reported on a quarterly basis, without reference to full- or part-time status of the employee or an actual date of hire or termination. A former student can hold more than one job in a quarter, either simultaneously or sequentially, which means that a rule must be adopted for handling multiple records for that person. The FETPIP management team has been an aggressive leader of, and participant with, colleagues in other states who have devised ways to grapple with these challenges.

- The FETPIP has a legislative mandate to collect and share information that must be used in making selected management decisions that affect Florida's public and private postsecondary colleges and universities. Over time this has led to a refinement of performance measurement and performance standards practices, which promises to be of substantial interest as congressional action on the array of pending reauthorizations and amendments mentioned earlier begins. The absence of a one-size-fits-all mentality in the FETPIP's activities will be of particular importance as evidence is sought to counter legislative and management fears of blunt instrument approaches to performance measurement.

- The FETPIP program has pioneered the successful pursuit of unit record database management without violating Federal or state confidentiality laws or public concerns about loss of privacy. This has been accomplished by ensuring that the laws themselves are understood and that all actions taken are in full compliance with these statutes and complementary regulatory provisions. In addition, the FETPIP's management team has taken a proactive stance in helping legislators, school officials, and the public-at-large to understand the legal provisions and FETPIP's features that protect individual rights to confidentiality.

A National, but Voluntary Approach

The FETPIP, its counterparts in other states that have a more limited interest in postsecondary information, and others in the employment and training, welfare-to-work, and related sectors, share a common interest in finding a practical way to gain access to information about employment and education events that occur beyond their own state's borders. Stevens and Duggan (1988) made the following recommendation:

> One possibility that Congress should consider for realizing this potential [of state unemployment insurance program administrative records] is the national archiving of the individual-level Wage Records now routinely purged in many of the states. The cost of computer processing and storage of millions of records was once prohibitive, but this is no longer the case. Current data processing technology far surpasses what was possible when the Wage Record programs came into existence [between 1938 and the later 1980s depending upon the referent state], and further advances in archiving methods can be expected. At the same time, routine procedures already have been developed to ensure protection of the anonymity of records and the privacy of individuals.[2]

The Federal government responded to this common need in 1995. The Unemployment Insurance Information Technology Support Center was created by the Unemployment Insurance Service in the Employment and Training Administration of the U.S. Department of Labor. A consortium of four

organizations collaborate in the management of this Center: (1) the State of Maryland, (2) Mitretek Systems, (3) Lockheed Martin Corporation, and (4) the University of Maryland. The overall mission of the Center is to serve as a laboratory and clearinghouse of up-to-date information about how state unemployment insurance units can take advantage of new information technologies to improve the quality of their services to clients and to control the cost of delivering these services.

Soon after the Center was established, the director of the U.S. Department of Labor's America's Labor Market Information System (ALMIS) entered into an agreement with the Center's leadership to design and pilot test an interstate sharing of wage record data to increase the value of state performance measurement activities like the FETPIP. A distributed database design was chosen so ownership and physical control of each state's administrative records would remain in the state; no creation of a single national database was envisioned at any time.

The Center's distributed Wage Record Interchange System (WRIS) concept is straightforward, although the logistics of implementing this conceptual design are not as simple. A requesting entity, such as Florida's FETPIP program, would enter into a formal data sharing agreement with the Information Technology Support Center. This agreement would specify the rules for participating as a sender and receiver of information. These stipulations would cover such issues as frequency and specific timing of anticipated requests, means of information transmittal, expected targets of the distributed inquiries, acceptable time lapse for responses, required formats of responses, financial considerations, and legal matters.

The challenges faced by the Center's design team for the distributed WRIS are readily apparent. What rules for eligibility to enter into a data sharing agreement will be established? Will each state be expected to designate a single entity to act as the administrative agent for all interested parties in that state, and if so, how will that entity be selected? Will data requests be bundled at the state level and then transmitted to the Center, which means that compromises must be negotiated within the state among the participating parties about the desired frequency of request and acceptable response time? Or will requests be accepted from multiple nodes within a state but perhaps then be held for specified periods to achieve economies in the subsequent bundling of multiple states' requests and transmittal through the distributed network? The choices among these options will have important system design consequences. What rules will be established for deciding which records are distributed to what nodes throughout the network of states? How will the design be affected by states that decline to participate and late entrants or early departures? What threshold level of participation, perhaps stratified in some way, must be achieved to justify establishment and maintenance of the distributed capability?

The WRIS is still in the design and pilot testing phase. NPEC member Marc Anderberg, and others in the U.S. Department of Labor's Dallas (TX) region, are cooperating in this pilot phase. Anderberg's participation complements his Consumer Report System consortium activities that have been under way as part of the Department's ALMIS initiative. This consortium's activities are led by the Texas State Occupational Information Coordinating Committee (SOICC). The Texas SOICC has adopted many features of Florida's FETPIP program in pursuing its own performance measurement and accountability activities. The NPEC's leadership is encouraged to establish an ongoing line of communication with the Information Technology Support Center as it continues its design and pilot testing.

A Bottom-Up and Voluntary Approach

Maryland's (then) Department of Economic and Employment Development entered into a data sharing agreement with a University of Maryland System institution in 1989 to establish an archive of Maryland wage records that could be used for the agency's and other research purposes. Since 1991 this archive has been managed by The Jacob France Center in the Merrick School of Business at the University of Baltimore. Until 1996 this was a unique arrangement between a state university and a state employment security agency.* Last year, Georgia State University entered into a similar agreement with Georgia's Department of Labor, and the data acquisition step is under way. Currently, the University of Missouri-Columbia is negotiating with Missouri's Department of Labor and Industry to create a similar university-based archive and research capability. Also last year, the University of California entered into a formal data sharing agreement with California's Employment Development Department. This will facilitate restricted access by authorized researchers to the Department's data, but no mention has been made of actually archiving unit record data at a University location.

The data sharing agreement that now exists between Maryland's Department of Labor, Licensing and Regulation and the University of Baltimore requires written authorization to be obtained for each proposed research use of the data archived in The Jacob France Center. Strict confidentiality protocols are followed. The Center receives funds from: (1) the Department of Labor, Licensing and Regulation, to support a scope-

*Wage records have been used by university faculty members, and others, for research purposes for more than 30 years. The chronology of this development and many citations to publications that used these data are found in: David W. Stevens (February 23, 1994), Research Uses of Wage Record Data: Implications for a National Wage Record Database, Washington, DC: Division of Occupational and Administrative Statistics, Bureau of Labor Statistics, U.S. Department of Labor, 12 pp.

of-work that is revised annually; (2) the Governor's Workforce Investment Board, to design and maintain an interagency performance measurement and accountability system; (3) the Andrew Mellon Foundation, in cooperation with Princeton University, to acquire and maintain a database of Baltimore City Public Schools and metro Baltimore public community college student records; and (4) the Sloan Foundation, in cooperation with The Urban Institute, MIT, and Jobs for the Future, Inc., to investigate changes in earnings inequality in recent years. Each of these funded projects requires written authorization from the data providers to conduct the proposed research. Separate data sharing agreements have been negotiated with the Baltimore City Public Schools, seven of the state's 16 public community colleges, and the University of Maryland System.

This bottom-up approach to establishing a database and using it for authorized purposes is very time-intensive. But, like the FETPIP example, the local ownership and control factor is extremely important. There may be no practical difference in the actuarial risk that is posed by transmitting data across long distances and then using it at a remote location, but education leaders express a strong preference for local (often meaning state) control of a database. This is the motivating force behind the distributed database design being developed by the Unemployment Insurance Information Technology Support Center.

A piece-it-together approach also inevitably suffers from a loss-of-value that: (1) accompanies the absence of any control over the data elements that are included in the databases provided by the cooperating institutions; (2) can be traced to quality differences among the databases that are eventually acquired; and (3) stems from time diverted to fund raising, the negotiation of data sharing agreements, and investment in learning the nuances of each institution's data. At the same time, there is a sense that value is gained from: (1) knowing more about the actual characteristics of each database and (2) engaging in quid pro quo arrangements to provide each institution with information that it values for its own purposes.

A Within-State, Self-Contained Approach

The University of North Carolina General Administration

The University of North Carolina (UNC) General Administration's approach, as it has been described by Working Group colleague Gary Barnes, illustrates both the strength and weakness of seeking a clearly defined limited objective through data sharing. The University Administration has conducted student

outcome research in cooperation with the North Carolina Employment Security Commission and other state organizations.

Among the advantages documented are: (1) relatively few interagency data sharing agreements had to be negotiated; (2) more was known at the outset about the quality of the student records because they were drawn from a single system, although not from just one institution; (3) the availability of a usable student identifier (social security number) was more likely, since this is the official designated student identifier, but accuracy was not ensured; and (4) use of the data could be controlled, so exposure to unauthorized use was not a problem.

The disadvantages, as described by Barnes, include: (1) disappointment with the limited findings that can be reported using the administrative records alone and (2) frustration that it is so difficult to build an effective coalition of those who are in a similar circumstance individually, but each of whom would benefit from a successful common assault on these constraints. The logical result is advocacy for an equally limited broadening of scope to cross state borders in search of valuable information, but without expanding the original intent of the data sharing activity. The distributed WRIS described earlier would accomplish the immediate goal sought, which is information about whether and where UNC "leavers" are working and how much they are earning, if the pilot effort is judged to be a success and enough states agree to cooperate in the distributed data sharing activity. The National Student Loan Clearinghouse may offer complementary access to continued pursuit of education activities by former UNC students.

University of Maryland System

The University of Maryland System Administration has developed an electronic transcript transmittal capability. The relatively new data sharing system is known as the Maryland Partnership in Electronic Data Interchange (EDI). The Administration's systems engineering unit designed the process to feature a central repository, or "post office," that accepts transcripts sent electronically by a previously authorized sending institution in Maryland to a designated, and also previously authorized, Maryland recipient. The security system includes a call-back feature to ensure that the request is actually coming from an approved node and a party who knows the proper passwords. Transcript information that has been sent and accepted using the call-back feature triggers a receipt detailing the date and time sent, and the date and time of receipt and person acknowledging that receipt.

This is a PC-based system in which necessary software and appropriate modem capabilities were provided by the University of Maryland System Administration to ensure a common quality of performance standard. The security features have resulted in no complaints of loss of confidentiality to date, since the mail box and call-back features are far more secure than the more costly, and subject to loss, mailing of paper transcripts that the system has replaced.

Interstate Data Sharing

There are numerous examples of ad hoc interstate sharing of outcomes information, but the approach adopted by Washington's State Board for Community and Technical Colleges (WSBCTC) was chosen for coverage here because it had been sustained longer before being affected by a unique event. A data sharing agreement was negotiated with Washington's Employment Security Department, but also with the state employment security agencies in other west coast states. For several years the WSBCTC was able to report the employment and earnings status of selected populations of former students that represented broader geographic coverage than other state-specific performance measurement systems.** The downside of this type of one-on-one negotiation of interstate data sharing agreements is illustrated by the fact that Oregon's Department of Employment sought a legal opinion from the state attorney general's office with respect to an intrastate data sharing initiative, which led to the termination of the then-active interstate data sharing agreement with Washington. Again, the proposed national distributed WRIS may help to resolve this issue, but optimism in this regard must rely on an expectation of each state's willingness and legal ability to participate. We are not aware of any reliable source of information about how many states can, and might be expected to, participate in this distributed database activity.

Multi-purpose Accountability Systems

There has been a substantial amount of activity of this type in recent years. Much of this state action was taken in anticipation of enactment of either the Careers Act or Job Training Consolidation Act by the last Congress, which did not happen. Among ongoing efforts that might be monitored in the future on the NPEC's

---

**Again, other states have carried out interstate record matching for a variety of designated populations, sometimes more than once. Washington's effort was early in this series and led to generally acknowledged added value to the overall effort, without jeopardizing confidentiality of student records.

behalf are California's Performance-Based Accountability Implementation Plan, which was approved by the California State Job Training Coordinating Council (SJTCC) on June 20, 1996, and released to the public 10 days later. This document was required by the California Senate's Bill 645, which had become law on January 1, 1996. The primary intent of this legislation is "to develop a tool to assess the accomplishments and measure the effectiveness of California's workforce preparation system."

The Special Committee for Performance-Based Accountability, which was established by the SJTCC has identified four customers for report cards that will be released in phases over the next five years. These four client groups are: (1) oversight entities, (2) state and local workforce preparation agencies, (3) individuals interested in jobs and careers, and (4) employers. The workforce preparation agencies designated in the law are defined by "shall" and "may" clauses. The law states that ". . . this system shall measure the performance of state and federally funded education and training programs. Programs to be measured may include programs in receipt of funds from [Federal and state laws are named]." The Committee's plan released in June 1996 then states that ". . . the Committee intends, for the purposes of the initial set of SB 645 report cards, that at least a representative sample of those who have participated in each of the programs listed in SB 645 will be included. To this end, the Committee asked each program to propose those it would prefer to include." The plan further states that "the SJTCC believes that SB 645 intends that all programs in the state whose purpose is to prepare any part of the workforce should ultimately be included in the report card system."

The California SJTCC Committee's activities in the next few years should be monitored on behalf of the NPEC because the successes and failures experienced in California will undoubtedly affect what is tried elsewhere through state legislation like SB 645. Some Working Group members have expressed alarm at the problems that might accompany coverage of traditional public and private 4-year academic, and 2-year transfer, postsecondary institutions and programs in performance-based accountability systems of this type. Other Working Group members have voiced similar concerns about the difficulty that will be encountered in issuing report cards on bundles of electronic course-taking that are assembled by individual students who have no intention of completing a traditional course of study at one or more institutions. [3]

The National Student Loan Clearinghouse[4]

The National Student Loan Clearinghouse is a nonprofit organization that was established to facilitate the process by which higher education institutions keep financial aid lenders and guaranty agencies apprised

of the enrollment status of recipients of student loans. The original purpose was to simplify and consolidate the processes used by individual educational institutions, loan organizations, and students in this regard. It works with all types of postsecondary entities which are eligible to participate in Title IV higher education programs. In the December 1996 NPEC Council meeting, a representative of the Clearinghouse reported that the nation's private and public postsecondary institutions were using the Clearinghouse's services on a voluntary basis to report on the enrollment status of nearly 9 million students. This represented approximately 60 percent of the enrollments at the time.

The Clearinghouse receives automated enrollment reports from participating institutions up to nine times during the academic year. Reports are sent within 30 days of the beginning of a term to certify loan payment deferments for enrolled students. The reports include data on all students enrolled in participating institutions. Specific data items include: name, social security number, institution code, enrollment status (full-time, half-time, less than half time, graduate student, withdrawal, etc.), status start date, term beginning and end dates, anticipated graduation date, and student address.

On April 20, 1997, the Clearinghouse staff announced that it was launching a new service built on the data collected from institutions. The service would be referred to as *TransferTrack*. The services are expected to be available to institutions participating in the Clearinghouse's basic services during the summer of 1997.

TransferTrack is designed to assist institutions in meeting requirements of the Federal *Student Right to Know and Campus Security Act* concerning the tracking of students who transfer from one institution to another. Institutions wishing to avail themselves of the service will provide a file to the Clearinghouse containing the identity of students they wish to track. The Clearinghouse staff will then search its data base for a specified time frame to determine if the identified students have reenrolled in another Clearinghouse-participating institution. A report will be generated and provided to institutions containing the records of students who transferred along with the name of the new institution and enrollment dates. A fee will be charged to users based on the volume of students to be tracked.

On March 20, 1997, representatives of the Texas Automated Student Follow-up System from the Texas SOICC and FETPIP met with representatives of the Clearinghouse regarding potential student follow-up applications. Representatives from both states were interested in piloting an effort to develop follow-up data on their states' students who enrolled in postsecondary programs beyond their state borders. While there was mutual interest in such activities, the Clearinghouse staff indicated that it could not participate in such an

activity until they determined how well the TransferTrack program would operate and how participating institutions would react to the service. Nevertheless, several options were discussed, including a trial program where the Clearinghouse would only identify the fact of an out-of-state enrollment with no institutional detail.

Representatives of the three organizations discussed the possibility of adding additional data items to the Clearinghouse database to increase its value as a follow-up and tracking tool. These included students' majors, program levels, graduation status, and grade-point average. There was additional discussion regarding the ability to track simultaneous enrollments and longitudinal transfers as well.

As a result of the meeting, it was agreed that a pilot test might be further developed after the Clearinghouse staff had the opportunity to assess interest and participation in the new TransferTrack program. This may be a possibility late in 1997 or early in 1998.

CHAPTER V
RECOMMENDATIONS

This paper reviews broad concepts affecting unit record and aggregate data collections. At the same time as there are increased demands on postsecondary institutions and governing bodies for outcome accountability, there are fewer resources with which to develop required information. Further, there are significant changes in the ways that students pursue postsecondary education that makes it difficult to define the accountable unit. Given these situations, institutions, states, and other responsible entities have recognized the growing importance of more efficient use of statistical resources by using record linkages -- that is, the exchange of unit-level data with other entities for statistical purposes. These exchanges may involve linking administrative records with other administrative records or survey data.

It is recommended that the Working Group on Unit Record Versus Aggregate Data continue its work into the next fiscal year at the direction of the NPEC Steering Committee. It is suggested that the next phase of the group's work focus particularly on potential linkages of unit record data to assist postsecondary entities in the development of data that will meet demands for accountability. This will require an analysis of what major unit record exchanges exist, what difficulties are inherent in the exchanges, and what key processes are required to ensure appropriate, efficient manipulation and reporting of data. The analyses should address privacy and data security, non-disclosure of individually identifiable attributes, and possible reidentification. They should also include how reporting issues are addressed, including non-coverage, duplicate records, and quantitative analysis. The uses of the data that result should be specified and illustrated. There are a number of parallel efforts in several states and federal agencies. It should be an important part of the Working Group's efforts to identify these and become familiar with how these issues are being addressed.

Endnotes

1. *Annual Report*, Tallahassee, FL: Florida Education and Training Placement Information Program (FETPIP), pp. 1-11, for a chronology of the 1984-1990 period and of the 1975-1984 activities that led to the creation of the FETPIP; and The Florida Education and Training Placement Information Program (April 16, 1996), Tallahassee, FL: Office of Workforce Education and Outcome Information Services, 5 pp. + tables and charts, which was Attachment 6 in a transmittal to NPEC Unit Record Versus Aggregate Data Working Group members last summer.

2. David W. Stevens and Paula Duggan (1988), *Labor Market Information: An Agenda for Congress*, Washington, D C: Northeast-Midwest Institute, p. 12.

3. Related documents that cover multi-purpose accountability systems include: John Baj et al. (1992), *Conclusions and Recommendations on the Development of the MACRO System*, DeKalb, IL: Center for Governmental Studies, Northern Illinois University; and, John Baj, Robert G. Sheets and Charles E. Trott (1994), *Building State Workforce Development Systems Based on Policy Coordination and Quality Assurance*, Washington, DC: Center for Policy Research, National Governors' Association; and, David W. Stevens (March 1, 1995), Baltimore, MD: Governor's Workforce Investment Board, 22 pp. + appendix.

4. This summary was added by Working Group Chairman Jay Pfeiffer following his visit and interviews with Vice President John Ward. Readers may wish to review the National Student Loan Clearinghouse document *TransferTrack: A New Service Developed to Provide Post-secondary Institutions with Information on Students Who Transfer Out to Other Schools*, National Student Loan Clearinghouse; Herndon, VA, May 1997.

United States
Department of Education
Washington, DC 20208–5652

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

Standard Mail (A)

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education

NCES 98–280